

A differential evolution algorithm with adaptive controlling weighted parameter for finite mixture model of some fire insurance data in Thailand

Somchit Boonthiem¹, Soontorn Boonta², Watcharin Klongdee^{1,*}

¹Risk and Insurance Research Group, Department of Mathematics, Faculty of Science, Khon Kaen University, Khon Kaen, 40002 Thailand

²Department of General Science, Faculty of Science and Engineering, Kasetsart University, Sakon Nakhon, 47000 Thailand

*Corresponding Author: kwatch@kku.ac.th

Received: 5 February 2017; Revised: 23 April 2017; Accepted: 24 April 2017; Available online: 1 August 2017

Abstract

We propose an adaptive differential evolution algorithm to estimate the parameter of finite mixture model. The finite mixture model separates two parts of parameter set. The first part consists of a parameter set of distributions that used for our experiment. The second part consists of a weighted parameter set of each component that are positive real value and the combination of them equal to 1. We perform an experiment to obtain an adaptive differential evolution algorithm for the finite mixture model by controlling the parameter set of the second part using a technique of weighting with constant such that parameter set of the second part satisfies condition of the finite mixture model. In this paper, we use an adaptive differential evolution algorithm to estimate the parameter of combination of nine distributions: Rayleigh distribution, logistic distribution, gamma distribution, Pareto distribution, log-logistic distribution, normal distribution, Weibull distribution, log-normal distribution, and exponential distribution for 47 claimed size data of fire insurance of an insurance company in Thailand which use K-S test statistic for the objective function. The results show that the best K-S test statistic value of finite mixture model which equal to 0.0535 less than the best single model which equal 0.0811. Therefore, the parameter set of the finite mixture model is 34.03 percent better.

Keywords: Weighted parameter; Finite mixture model; Fire Insurance; DE algorithm

1. Introduction

Standard Differential Evolution

Differential evolution (DE) is an optimization method that generates variations of the parameter vectors by iteration to improve a candidate solution with concern to a given objective function. The general strategy of DE has four stages as shown in Fig. 1 which is proposed by Storn and Price [1].

The first stage is a construction of the initial population solution called Initialization. Set NP , a number of population for each generation m , and

$$T_i^{(m)} = (t^{(m)}(i,1), t^{(m)}(i,2), \dots, t^{(m)}(i,d))$$

which is the d -dimensional initial parameter vector by random a number belong to $(0,1)$ called the i^{th} target vector where $i = 1, 2, \dots, NP$. This stage must be computed the objective function in the final step.

The second stage is called the mutation which is the generating of new parameters by adding the weighted difference between two population vectors. For each target vector, the randomly chosen indexes $r_1, r_2, r_3 \in \{1, 2, \dots, NP\}$ are mutually different integers, and $F > 0$. A mutant vector

$$V_i^{(m)} = (v^{(m)}(i,1), v^{(m)}(i,2), \dots, v^{(m)}(i,d))$$

is generated in agreement with

$$v^{(m)}(i,j) = t^{(m)}(r_1,j) + F(t^{(m)}(r_2,j) - t^{(m)}(r_3,j)) \quad (1)$$

which is also known as **DE/rand/1** or other variants of the mutation as proposed in [2 – 6].

The third stage is called the crossover (or recombination) which is mutation between target vector's entries and mutant vector's entries. The new vector is called trial and denoted by

$$U_i^{(m)} = (u^{(m)}(i,1), u^{(m)}(i,2), \dots, u^{(m)}(i,d))$$

The crossover process is in the form

$$u^{(m)}(i,j) = \begin{cases} v^{(m)}(i,j) & \text{if } \text{randun}(i,j) \leq CR \text{ or } j = \text{randin}(i,j) \\ t^{(m)}(i,j) & \text{elsewhere,} \end{cases} \quad (2)$$

where $\text{randun}(i,j)$ is the j^{th} evaluation of a uniform random number generator with outcome belong to $[0,1]$ for all i . CR is the crossover constant belong to $[0,1]$ which has to be determined by the user and $\text{randin}(i,j)$ is a randomly chosen index belong to $\{1, 2, \dots, d\}$ which ensures that $U_i^{(m)}$ gets at least one parameter from $V_i^{(m)}$.

The final stage is called the selection which is the consideration under condition that if the trial vector yields a better objective function value than the target vector, then the target vector is replaced by the trial vector in the following generation. Otherwise, the old target vector is retained.

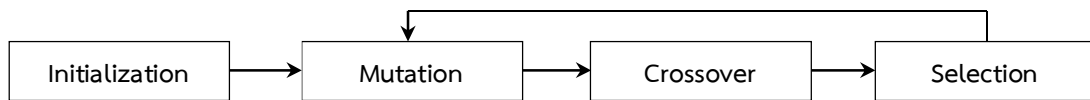


Fig. 1 Main stages of differential evolution algorithm

Finite Mixture Model (FMM)

Let X_1, X_2, \dots, X_n be random variables with distribution $F(x; \theta_1), F(x; \theta_2), \dots, F(x; \theta_n)$, respectively.

p_1, p_2, \dots, p_n are positive weighted parameters such that $\sum_{i=1}^n p_i = p_1 + p_2 + \dots + p_n = 1$. Next, we denote

$\Theta = (\theta_1, \theta_2, \dots, \theta_n)$ and the distribution function satisfies

$$F(x; \Theta) = \sum_{i=1}^n p_i F_i(x; \theta_i) \tag{3}$$

that is called the FMM. If $n = 1$, the distribution is called the single model. Throughout this paper, we consider the probability density function (p.d.f.) and the cumulative distribution function (c.d.f.) of nine distributions as shown as below.

Rayleigh distribution

The probability density function of the Rayleigh distribution has a positive scale parameter β_1 . Set $\theta_1 = \beta_1$, thus the p.d.f. and c.d.f. are shown in equations (4) and (5), respectively

$$f_1(x; \theta_1) = \frac{x}{\beta_1^2} \exp\left(-\frac{x^2}{2\beta_1^2}\right), x \geq 0, \tag{4}$$

$$F_1(x; \theta_1) = 1 - \exp\left(-\frac{x^2}{2\beta_1^2}\right), x \geq 0. \tag{5}$$

Logistic distribution

The logistic distribution has a positive scale parameter β_2 . Set $\theta_2 = \beta_2$, thus the p.d.f. and c.d.f. are shown in equations (6) and (7), respectively

$$f_2(x; \theta_2) = \frac{1}{\beta_2} \frac{\exp\left(-\frac{x}{\beta_2}\right)}{\left(1 + \exp\left(-\frac{x}{\beta_2}\right)\right)^2}, x \in \mathbb{R}, \tag{6}$$

$$F_2(x; \theta_2) = \frac{1}{1 + \exp\left(-\frac{x}{\beta_2}\right)}, x \in \mathbb{R}. \tag{7}$$

Gamma distribution

The gamma distribution has a positive shape parameter β_3 and a positive scale parameter β_4 . Set $\theta_3 = (\beta_3, \beta_4)$, thus the p.d.f. and c.d.f. are shown in equations (8) and (9), respectively

$$f_3(x; \theta_3) = \frac{1}{\beta_4^{\beta_3} \Gamma(\beta_3)} x^{\beta_3-1} \exp\left(-\frac{x}{\beta_4}\right), x \geq 0, \tag{8}$$

$$F_3(x; \theta_3) = \frac{\lambda\left(\beta_3, \frac{x}{\beta_4}\right)}{\Gamma(\beta_3)}, x \geq 0, \tag{9}$$

where Γ is the gamma function and λ is the lower incomplete gamma function.

Pareto distribution

The Pareto distribution has a positive shape parameter β_5 and a positive scale parameter β_6 . $\theta_4 = (\beta_5, \beta_6)$, thus the p.d.f. and c.d.f. are shown in equations (10) and (11), respectively

$$f_4(x; \theta_4) = \frac{\beta_5 \beta_6^{\beta_5}}{x^{\beta_5+1}}, x \geq \beta_6, \tag{10}$$

$$F_4(x; \theta_4) = 1 - \left(\frac{\beta_6}{x}\right)^{\beta_5}, x \geq \beta_6. \tag{11}$$

Log-logistic distribution

The log-logistic distribution has a real location parameter β_7 and a positive real scale parameter β_8 . Set $\theta_5 = (\beta_7, \beta_8)$, thus the p.d.f. and c.d.f. are shown in equations (12) and (13), respectively

$$f_5(x; \theta_5) = \frac{1}{\beta_8 x} \frac{\exp(z)}{(1 + \exp(z))^2}, x \geq 0, \tag{12}$$

$$F_5(x; \theta_5) = \left(1 + \left(\frac{x}{\beta_7}\right)^{-\beta_8}\right)^{-1}, x \geq 0. \tag{13}$$

where $z = \frac{\ln(x) - \beta_7}{\beta_8}$.

Normal distribution

The normal distribution (or Gaussian distribution) is the most important distribution in statistics which has a real location parameter β_9 and a positive real scale parameter β_{10} . Set $\theta_6 = (\beta_9, \beta_{10})$, thus the p.d.f. and c.d.f. are shown in equations (14) and (15), respectively

$$f_6(x; \theta_6) = \frac{1}{\beta_{10} \sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{x - \beta_9}{\beta_{10}}\right)^2\right), x \in \mathbb{R}, \tag{14}$$

$$F_6(x; \theta_6) = \frac{1}{2} \left(1 + \operatorname{erf}\left(\frac{x - \beta_9}{\beta_{10} \sqrt{2}}\right)\right), x \in \mathbb{R}, \tag{15}$$

where $\text{erf}(t) = \frac{2}{\sqrt{\pi}} \int_0^t \exp(-x^2) dx$.

Weibull distribution

The Weibull distribution has a positive real shape parameter β_{11} and a positive real scale parameter β_{12} . Set $\theta_7 = (\beta_{11}, \beta_{12})$, thus the p.d.f. and c.d.f. are shown in equations (16) and (17), respectively

$$f_7(x; \theta_7) = \frac{\beta_{11} x^{\beta_{11}-1}}{\beta_{12}^{\beta_{11}}} \exp\left(-\left(\frac{x}{\beta_{12}}\right)^{\beta_{11}}\right), x \geq 0, \tag{16}$$

$$F_7(x; \theta_7) = 1 - \exp\left(-\left(\frac{x}{\beta_{12}}\right)^{\beta_{11}}\right), x \geq 0. \tag{17}$$

Log-normal distribution

The log-normal distribution has a real location parameter β_{13} and a positive real scale parameter β_{14} . Set $\theta_8 = (\beta_{13}, \beta_{14})$, thus the p.d.f. and c.d.f. are shown in equations (18) and (19), respectively

$$f_8(x; \theta_8) = \frac{1}{\beta_{14} x \sqrt{2\pi}} \exp\left(-\frac{(\ln x - \beta_{13})^2}{2\beta_{14}^2}\right), x > 0, \tag{18}$$

$$F_8(x; \theta_8) = \frac{1}{2} + \frac{1}{2} \text{erf}\left(\frac{\ln x - \beta_{13}}{\sqrt{2}\beta_{14}}\right), x > 0. \tag{19}$$

Exponential distribution

The exponential distribution has a positive scale parameter β_{15} . Set $\theta_{15} = \beta_{15}$, thus the p.d.f. and c.d.f. are shown in equations (20) and (21), respectively

$$f_9(x; \theta_9) = \frac{1}{\beta_{15}} \exp\left(-\frac{x}{\beta_{15}}\right), x \geq 0, \tag{20}$$

$$F_9(x; \theta_9) = 1 - \exp\left(-\frac{x}{\beta_{15}}\right), x \geq 0. \tag{21}$$

Nowadays, parameter estimation for mixture model is interesting in various fields such as loss model of insurance, financial indicators model, extreme climatological conditions model, etc. A claim size data of fire insurance is extreme value distribution which is interested for insurers in order to management risk occurrence through a loss model. There are many researchers who study loss model of fire insurance by using some claim sizes. In 2016, Miljkovic and Grün [7] proposed an alternative approach for flexible modeling of heavy tailed, such as the well-known data set on Danish Fire losses with the EM algorithm. The result of their experiment shown that the mixture model are compared to the composite Weibull models considered in recent literature as the best models for modeling Danish Fire insurance losses. In 2016, Boonthiem, *et al.* [8] used some fire insurance data of claim size in Thailand to estimate the parameter of nine distributions as mentioned above with minimum

K-S estimator using grid search technique which obtained the K-S test statistic value for the nine distributions as 0.12698, 0.44620, 0.10392, 0.08110, 0.11193, 0.12797, 0.14659, 0.11887, and 0.30005, respectively. Therefore, we are interested in the minimizing K-S test statistic value of a finite mixture model of nine distributions using DE algorithm.

2. Materials and Methods

Data and objective function

In this paper, we use some data of claim severities of fire insurance in Thailand from 2000 to 2004 provided by the Thai Reinsurance Public Co., Ltd. as shown in Table 1 and Fig. 2.

Table 1 The claim size data of the fire insurance in Thailand (million)

Times	Claim size	Times	Claim size	Times	Claim size
1	35.5	17	26.7	33	69.9
2	26.4	18	32.4	34	122.7
3	64.9	19	76.5	35	158.9
4	127.3	20	33.2	36	33.1
5	57.7	21	25.7	37	60
6	21.8	22	60.2	38	104.3
7	67.3	23	132.2	39	29.2
8	48.5	24	20.9	40	63.1
9	23.6	25	65.8	41	90
10	22.3	26	55.3	42	27.2
11	84.6	27	33	43	22.4
12	21.4	28	22.1	44	27.5
13	51.5	29	24.2	45	57.2
14	20.7	30	44.4	46	34.2
15	40.1	31	20.4	47	53.2
16	29.3	32	30.8		

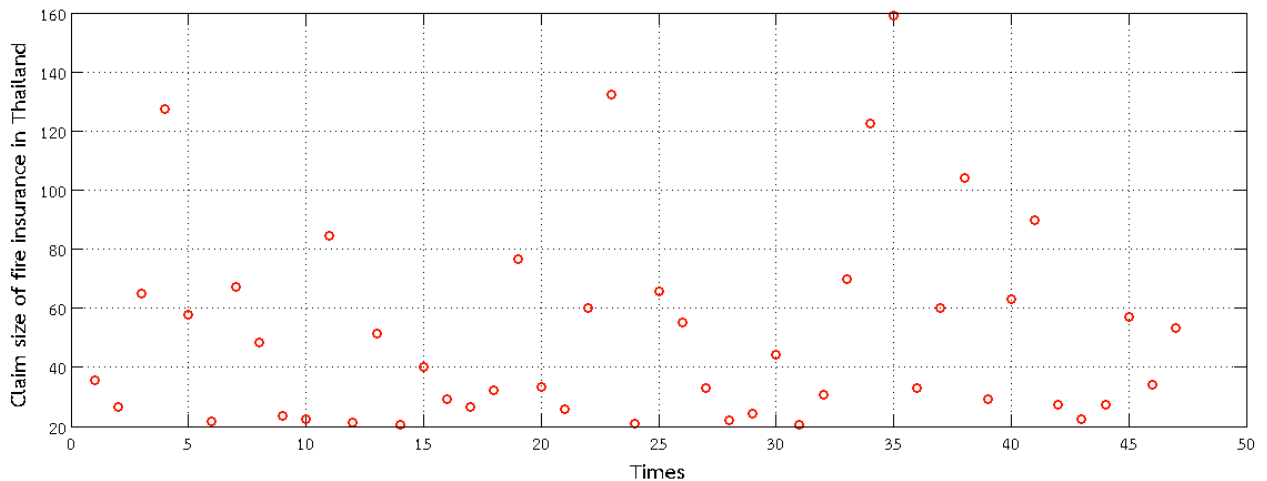


Fig. 2 Claim size data of fire insurance in Thailand

Let x_1, x_2, \dots, x_N such that $x_1 < x_2 < \dots < x_N$ be a sample of N independent and identically distributed observations of a real-valued one-dimensional random variable X that has parameter vector Θ . The cumulative distribution function (CDF) of X is denoted by $F_x(x_i; \Theta)$. The K-S test statistic value of $F_x(x_i; \Theta)$ is obtained by

$$KS(\Theta; X) = \max_{1 \leq i \leq N} \left| F_x(x_i; \Theta) - \frac{i}{N} \right|. \tag{22}$$

The K-S test statistic is a goodness of fit test that uses to measure the deviation between the predicted data using theoretical probability function and the observed data that is used for our objective function.

Adaptive DE Algorithm with Controlling the Weighted Parameter

In this subsection, we propose a differential evolution algorithm to minimize K-S test statistic value of finite mixture model as mentioned in the previous section. Generally, algorithm which is used for estimating the finite mixture model is expectation and maximization (EM) method. In addition, other methods like Bayesian technique and maximum likelihood estimation can stabilize these estimates. Our algorithm proposes four stages similar to Storn and Price mentioned in subsection *Standard Differential Evolution*.

Stage of initialization, given an objective function with correspond to equation (22), and then fixed the scaling factor $F = 0.4$, the crossover rate $CR = 0.8$ and the number of population $NP = 50$. Let $\Theta^{(0)} = (\theta^{(0)}, p^{(0)})$ be an initial parameter vector where $\theta^{(0)} = (\beta_1^{(0)}, \beta_2^{(0)}, \dots, \beta_{15}^{(0)})$ and $p^{(0)} = (p_1^{(0)}, p_2^{(0)}, \dots, p_9^{(0)})$ are the target vector which are uniformly randomized such that $0 < \theta^{(0)} < 1, 0 < p^{(0)} < 1$. Next, we consider the condition of controlling the weighted parameter $p^{(0)}$, $p_j^{(1)} = \frac{p_j^{(0)}}{\sum_{j=1}^9 p_j^{(0)}}$. if $\left| \sum_{j=1}^9 p_j^{(0)} - 1 \right| < 0.001$ where $j = 1, 2, \dots, 9$. Lastly, we must compute the objective function of each i .

Stage of mutation, determine a bounded parameter set using initial scale parameters of distributions in [8]. Then, randomly distinct integer numbers r_1, r_2 and r_3 belong to $\{1, 2, \dots, NP\}$. Next, compute value of mutant vector that satisfies equation (1). In this stage, the new weighted parameter $p_j^{(2)} = \frac{p_j^{(1)}}{\sum_{j=1}^9 p_j^{(1)}}$ must be controlled identically with the condition of initialization stage. The final stage, compute the objective function of each i .

Stage of crossover, let R be a uniformly random number in $(0, 1)$ which has $NP \times$ number of parameter dimension. Next, consider the mutant vector in the previous stage to recombine using conditions of equation (2).

Next, controlling the new weighted parameter $p_j^{(3)} = \frac{p_j^{(2)}}{\sum_{j=1}^9 p_j^{(2)}}$ is identical the stage of initialization. Additional,

compute the objective function for each i .

Stage of selection, the entries of new target vector are selected by comparing the minimum K-S test statistic value of target vectors with the trial vector, i.e., for each i , if $KS(\Theta; X)$ of $T^{(m)}(i, j)$ is less than $KS(\Theta; X)$ of $U^{(m+1)}(i, j)$, then the parameter in the i^{th} row of the best parameter vector are obtained $B^{(m)}(i, j)$. Otherwise, the parameter in the i^{th} row of the best parameter vector are obtained $U^{(m+1)}(i, j)$. After that,

control the new weighted parameter $p_j^{(4)} = \frac{p_j^{(3)}}{\sum_{j=1}^9 p_j^{(3)}}$ which is identical to the stage of initialization and compute

the objective function for each i . Set $B^{(m)} = [B^{(m)}(1, j), B^{(m)}(2, j), \dots, B^{(m)}(NP, j)]^T$. Choose the best parameter set of $B^{(m)}$ from minimum K-S test statistic value denoted $B^{(m)}(best, j)$.

Next, we repeat this process for 500 and 1000 generations of four stages until the best parameter set satisfies the stopping condition which is

$$\left| B^{(m+1)}(best, j) - B^{(m)}(best, j) \right| < 0.5 \times 10^{-10} \tag{23}$$

such that continuously counting for 50 and 100 generations.

3. Results and Discussion

Our purpose of this paper is to minimize the K-S test statistic value of a mixture model. The experiment is followed the adaptive DE algorithm with controlling the weighted parameter. The algorithm uses the criteria in an inequality (23) to stop the process for 500 and 1000 generations by continuously counting 50 and 100 generations. The results for 500 and 1000 generations by continuously counting 50 generations are shown in Table 2 and the K-S test statistic value is shown in Fig. 3 and Fig. 4. The results for 500 and 1000 generations by continuously counting 100 generations are shown in Table 3 and the K-S test statistic value is shown in Fig. 5 and Fig. 6. The comparison between the best K-S test statistic value of a single model [9] and the best K-S test statistic value of a finite mixture model show that the K-S test statistic value of the mixture model is better than the K-S test statistic value of the single model.

Table 2 Parameters of distributions of experiment for $F = 0.4$, $CR = 0.8$, $NP = 50$ for 500 and 1000 generations by continuously counting 50 generations

Distributions	Parameters for 500 generations				Parameters for 1000 generations			
	Scale	Location	Shape	Weighe	Scale	Locatio	Shape	Weighed
Rayleigh	47.327175	-	-	0.002254	45.981928	-	-	0.000995
Logistic	21.457491	-	-	0.001679	20.251788	-	-	0.004957
Gamma	29.209236	-	1.513830	0.368970	26.264258	-	1.726550	0.405938
Pareto	25.876277	-	1.017927	0.553112	25.744983	-	1.012177	0.459329
Log-logistic	1.561592	7.473530	-	0.006457	5.831469	21.197970	-	0.067216
Normal	36.922146	6.514544	-	0.019067	36.186541	2.341783	-	0.003382
Weibull	3.910071	-	3.537471	0.040926	3.757824	-	6.601435	0.006727
Log normal	2.819982	1.455072	-	0.006314	2.779355	1.254609	-	0.001476
Exponential	54.285534	-	-	0.001221	55.465405	-	-	0.049980
$KS(\Theta; X)$	0.054410				0.053526			

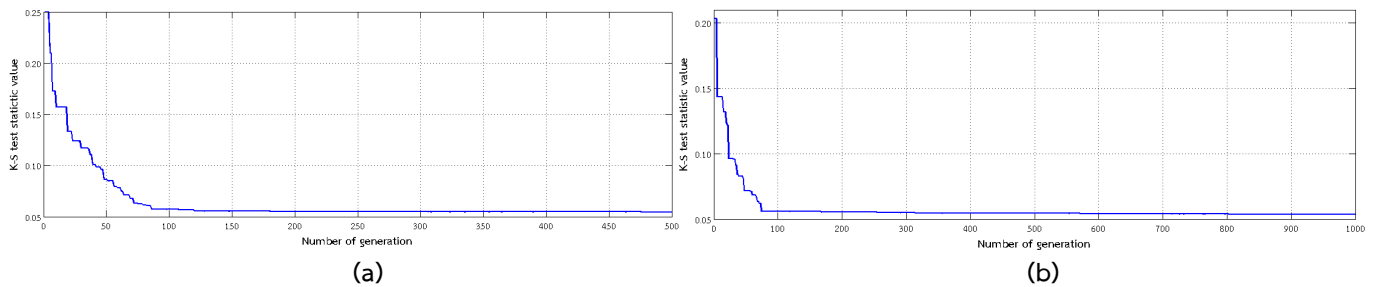


Fig. 3 The K-S test statistic for $F = 0.4$, $CR = 0.8$, $NP = 50$, stopping at 500 (a) and 1000 (b) generations by continuously counting 50 generations

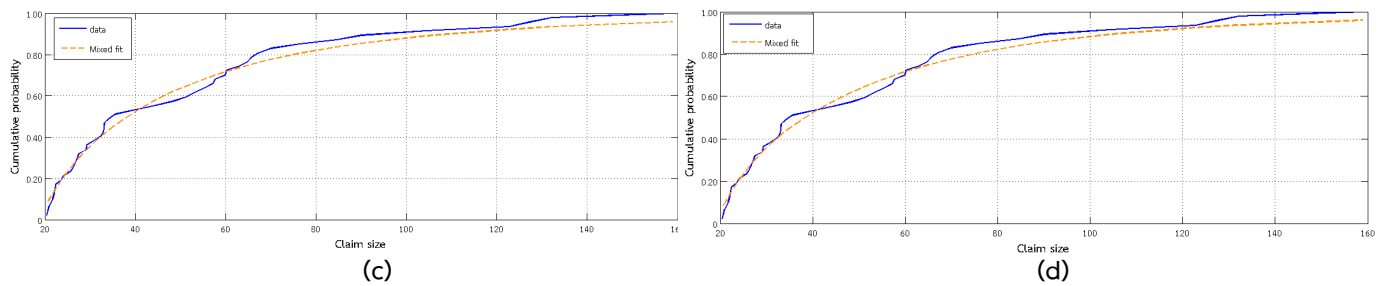
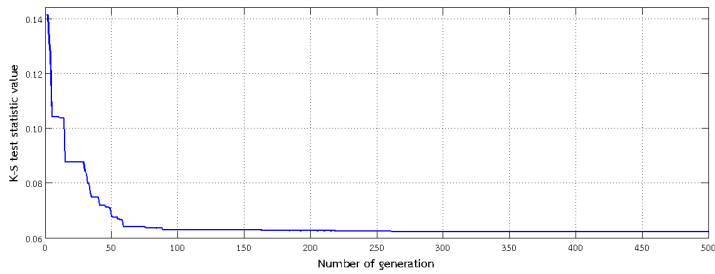


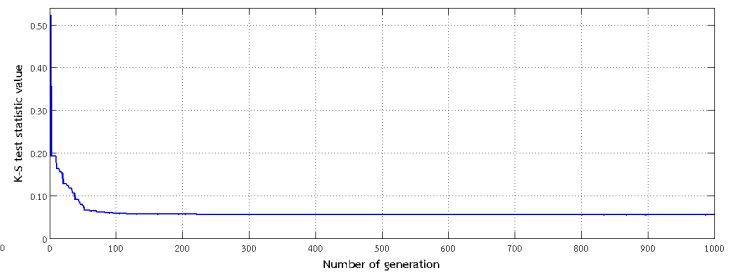
Fig. 4 The mixture model fitted for $F = 0.4$, $CR = 0.8$, $NP = 50$, stopping at 500 (c) and 1000 (d) generations by continuously counting 50 generations

Table 3 Parameters of distributions of experiment for $F = 0.4$, $CR = 0.8$, $NP = 50$ for 500 and 1000 generations by continuously counting 100 generations

Distributions	Parameters for 500 generations				Parameters for 1000 generations			
	Scale	Location	Shape	Weighe	Scale	Locatio	Shape	Weighed
Rayleigh	43.798346	-	-	0.017924	44.112308	-	-	0.001440
Logistic	21.059049	-	-	0.019634	23.313091	-	-	0.003690
Gamma	23.733079	-	1.516681	0.365633	24.755184	-	1.506175	0.345379
Pareto	25.048360	-	0.943284	0.447548	25.963926	-	0.976590	0.468250
Log-logistic	3.271585	0.880754	-	0.021951	5.500672	13.780891	-	0.018437
Normal	35.528685	3.077117	-	0.018495	38.321589	0.206249	-	0.034782
Weibull	3.036850	-	5.903329	0.044977	6.613765	-	6.954007	0.018592
Log normal	5.030176	2.309865	-	0.000384	2.527939	0.905190	-	0.024732
Exponential	54.373795	-	-	0.063454	53.119823	-	-	0.084698
$KS(\Theta; X)$	0.055471				0.055174			

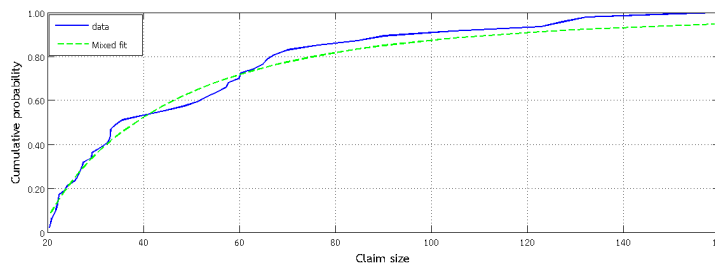


(e)

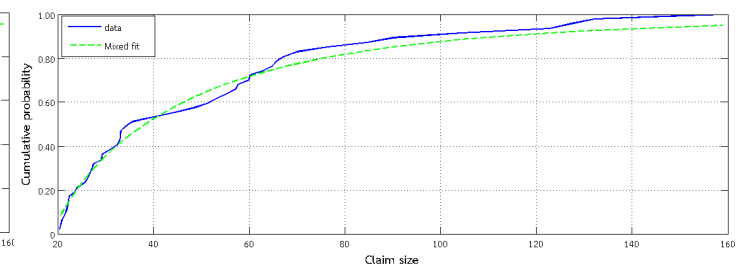


(f)

Fig. 5 The K-S test statistic for $F = 0.4$, $CR = 0.8$, $NP = 50$, stopping at 500 (e) and 1000 (f) generations by continuously counting 100 generations



(g)



(h)

Fig. 6 The mixture model fitted for $F = 0.4$, $CR = 0.8$, $NP = 50$, stopping at 500 (g) and 1000 (h) generations by continuously counting 100 generations

4. Conclusion and Suggestions

DE algorithm for the finite mixture model of nine distributions: Rayleigh distribution, logistic distribution, gamma distribution, Pareto distribution, log-logistic distribution, normal distribution, Weibull distribution, log-normal distribution, and exponential distribution is suitably with our data set of experiment by setting scaling factor $F = 0.4$, crossover rate $CR = 0.8$, number of population $NP = 50$, and use the criteria in the inequality

(23) in order to stop the process. The results show that the best K-S test statistic value of finite mixture model which equal to 0.0535 less than the best single model which equal 0.0811. Consequently, the parameter set of the finite mixture model is 34.03 percent better.

In the future, the applications of DE algorithm have usefulness for parameter estimation of the finite mixture model and researchers may improve the proposed algorithm to be a better algorithm. The model's applications can be used to compute the minimum initial capital as presented by Khotama, et al. [8]. Moreover, efficient and credible generating of the data set is the most important.

5. Acknowledgement

This work is supported by Risk and Insurance Research Group, Department of Mathematics, Faculty of Science, Khon Kaen University, and Mathematics and Statistics Program, Faculty of Science and Technology, Sakon Nakhon Rajabhat University, Thailand.

6. References

- [1] R. Storn, K. Price, Differential Evolution – A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces, *J. Global Optim.* 11 (1997) 341 – 359.
- [2] M. Ali, M. Pant, A. Abraham, Simplex Differential Evolution, *Acta Polytech Hung.* 6(5) (2009) 95 – 115.
- [3] V.V. de Melo, L.C.G. Carosio, Investigating Multi-View Differential Evolution for solving constrained engineering design problems, *Expert Syst Appl.* 40(2013) 3370 – 3377.
- [4] E. Mininno, F. Neri, F. Cupertino, D. Naso, Compact Differential Evolution, *IEEE Trans. Evol. Comput.* 15(1) (2011) 32 – 54.
- [5] F. Neri, V. Tirronen, Recent advances in differential evolution: a survey and experimental Analysis, *Artif Intell Rev.* 33(1-2) (2010) 61 – 106.
- [6] M. Weber, F. Neri, V. Tirronen, A study on scale factor in distributed differential evolution, *Information Sciences.* 181(2011) 2488 – 2511.
- [7] T. Miljkovic, B. Grün, Modeling loss data using mixtures of distributions, *Insur Math Econ.* 70(2016) 387 – 396.
- [8] S. Boonthiem, S. Khotama, S. Sakha, W. Klongdee, Minimum K-S estimator using PH-transform technique, *SSSTJ.* 3(2) (2016) 11 – 16.
- [9] S. Khotama, T. Thongjunthug, K. Sangaroon, W. Klongdee, On Approximating the Minimum Initial Capital of Fire Insurance with the Finite-time Ruin Probability using a Simulation Approach, *KKU Res. J.* 20(3) (2015) 267 – 271.